JOURNAL OF GEOSCIENCE EDUCATION **59**, 41–45 (2011)

Group-Examination Improves Learning for Low-Achieving Students

G. L. Macpherson,[1,a)] Young-Jin Lee,[2] and Don Steeples[1]

ABSTRACT

An introductory geology class that satisfies a liberal arts distribution requirement was used to investigate the benefits of allowing discussion during assessments. For three term examinations, students completed short- to medium-length essay tests individually (individual examination) and then again as part of an assigned group of four to five students (group examination). The comprehensive final examination for the course was multiple-choice and true-false questions, with 75% of the questions covering material on the term examinations and 25% of the questions covering material not tested previously. Students generally favored the group examinations, both midway through the course and at the end of the course, but final examination results were mixed. Those whose scores increased the most in the group examinations tended to have higher percentage correct on both previously tested and new material on the final examination. Those whose group-examination scores were not much better than their individual scores performed at a level similar to or slightly worse than their performance on the term examinations. This suggests that low-achieving students benefit the most from the group examinations. Using the group-examination format in a large class will require a clear rubric for grading by multiple graders, but is expected to result in higher success for low achievers. © *2011 National Association of Geoscience Teachers*. [DOI: 10.5408/1.3543930]

INTRODUCTION

Undoubtedly, the most popular instructional strategy used in entry-level science (including geoscience) courses in higher education institutions is lecturing. Although lecturing is a highly efficient means to deliver a large amount of information to a large number of students, it may not be able to effectively engage students in active learning because of the passive nature of its instructional processes. According to a concept called *constructivism*, in order to truly learn something new, students must actively "construct" their own understanding for themselves and connect this new understanding to what they already know while discarding any misconceptions that they may have had (Tobin, 1993).

One measure of learning is performance on assessments. In three sections of a large lecture class serving a population seeking to fulfill an earth science distribution requirement (Earthquakes and Natural Disasters), a student's score on the second term examination successfully predicts the student's final grade in the class (Fig. 1; $r = 0.69 \pm 0.01r$), even though that term examination is weighted as only 28%–33% of the final score. Furthermore, for more than ~85%–90% of the students, the letter grade (A, B, C, D, F) assigned for the course was lower or equal to the letter grade earned on the second midterm examination (Fig. 2). To see if there was a way to improve learning, engagement, and performance in this type of class, we used an intensive, small-enrollment section of this same course (class size < 20) taught in a summer session, where student focus was directed on only this course. In contrast, the large enrollment class (class size > 600 individuals; Figs. 1 and 2) was a more traditional lecture class taught during regular academic-year semesters. This paper presents the results of using a cooperative learning strategy in the small-enrollment class. The outcomes will be useful in choosing successful aspects of the strategy and modifying it for use in a large-enrollment class, in the future.

Researchers have been trying to develop an effective learning strategy, such as cooperative learning, in order to engage students in more active learning. Cooperative learning is one of the instructional strategies designed to implement constructivism by requiring students to communicate and cooperate with each other and thus become more actively involved in accomplishing a common learning task. Cooperative learning could facilitate students' problem solving (and thereby result in long-term retention of the learned knowledge) because during collaboration cognitive conflicts could arise, inadequate reasoning could be corrected, and therefore enriched understanding could be achieved (Ben-Ari and Kedem-Friedrich, 2000; Blumenfeld *et al.*, 1996; Pear and Crone-Todd, 2002; Vygotsky, 1978). Education researchers have been developing several teaching methods based on cooperative learning, such as peer instruction, think-pair-share, the jigsaw technique, and the tiered exam.

Peer instruction involves having students individually solve a multiple-choice concept question where each choice in the concept question reflects common misconceptions, and then solve the same question in a pairs or in a small group again (Mazur, 1997). This technique has been successfully used in many Science, Technology, Engineering and Mathematics (STEM) disciplines such as physics (Crouch and Mazur, 2001; Fagen *at al.*, 2002), chemistry (Kovac, 1999), and geoscience (McConnell *et al.*, 2006). "Think-pair-share" is similar to peer instruction in that the instructor poses a question or problem and asks all students to discuss their answers with their neighbors to share views or to reach consensus (Macdonald and Korinek, 1995; Yuretich *et al.*, 2001), but the question or problem used in think-pair-share may not be limited to research-based multiple-choice concept questions typically used in the peer instruction method. While peer instruction and think-pair-share do not impose a specific limitation on how students form a pair or a small

Received 7 August 2009; accepted 30 November 2010; published online 22 February 2011.

[1]Department of Geology, University of Kansas, Lawrence, Kansas 66045-7613, USA
[2]Department of Educational Leadership and Policy Studies, University of Kansas, Lawrence, Kansas 66045-7613, USA
[a)]Author to whom correspondence should be addressed. Electronic mail: glmac@ku.edu.

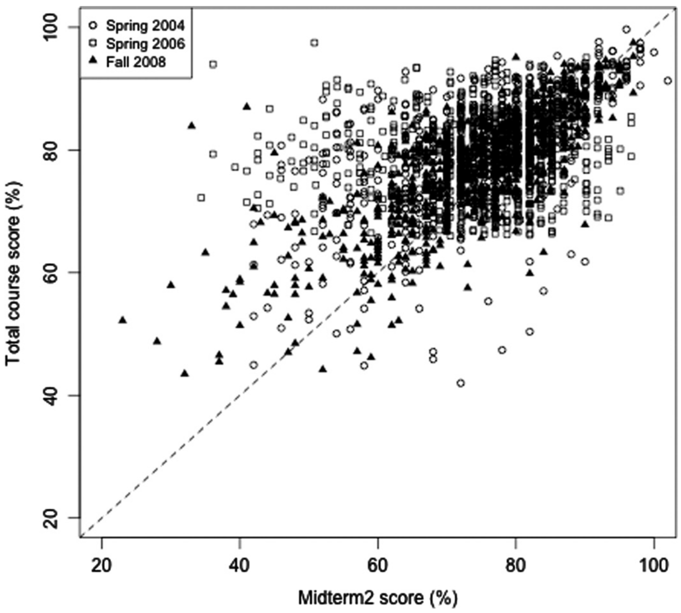

FIGURE 1: Comparison of percentage-correct scores on second midterm examination in three semesters of large-enrollment sections to percentage score for the entire semester for the same students. The correlation is strong (r = 0.69 ± 0.01r), suggesting that the score on the second midterm examination is a good predictor of a student's score for the entire semester. The first midterm examination correlation to the entire-semester score (not shown) is not so strong (r = 0.54 ± 0.01) as the second midterm correlation shown here, and the third midterm examination correlation to the entire-semester score is about the same (r = 0.70 ± 0.01r) as the second midterm correlation shown here.

group for their discussion, the jigsaw technique requires students to create a "team" to maximize their learning outcomes from the cooperative learning processes. In this approach, teams of students are assigned to investigate *different* aspects of the same problem. Once teams have completed their assignments, members of each team then disperse

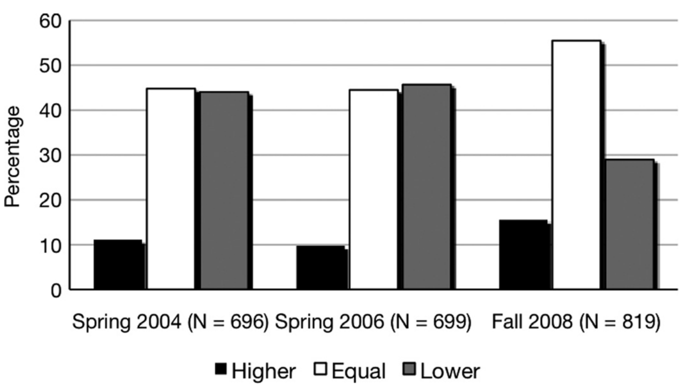

FIGURE 2: The percentage of students receiving a higher final letter grade for the semester than the letter grade-equivalent to the second midterm exam is less than about 15% for all three sections of large-enrollment classes analyzed. These results suggest that students do not tend to improve their grade over a semester, but maintain (or even lower) their grade.

among new groups and teach group members from other teams about what they have learned (Bykert-Kauffman, 1995; Tewksbury, 1995). Finally, the tiered exam, which is also called the pyramid exam or the two-stage exam, is an instructional approach trying to transform an examination into more meaningful learning experiences. The tiered exam typically consists of an initial traditional close-book examination, followed by a second open-book, open-note examination during which peer teaching is encouraged (Tewksbury and Macdonald, 2005; Sheldon 2005).

The objective of this study was to investigate the educational benefit of a new cooperative learning approach combining the jigsaw technique and tiered exam in a small-enrollment course. In particular, we were interested in finding out whether this new cooperative learning approach applied during the midterm examinations could result in improved performance on the final examination where collaboration among students was NOT permitted.

METHOD

In order to investigate the usefulness of our new cooperative learning method, which applies the jigsaw technique in the context of the tiered exam, we used an intensive, small-enrollment class (Earthquakes and Natural Disasters, $N = 19$) that fulfills liberal arts requirement. The large enrollment class sections of this course, where learning does not seem to improve during a term (Fig. 1), differ in some ways from the small enrollment section. We are not attempting to contrast results on these two different types of classes directly, but note that the small-enrollment class was dominated by seniors rather than by freshman or sophomores in the large-enrollment sections, and the population in the small-enrollment section was less strongly dominated by liberal arts majors than in the large-enrollment sections. However, both small- and large-enrollment sections included those studying a wide variety of disciplines, with most students pursuing degrees in liberal arts. In addition, both small- and large-enrollment classes served the purpose of fulfilling a distribution requirement for graduation, used the same textbook, and provided instruction in similar topics. The final grade percentages earned by students in both small and large sections were similar, as were the ranges of final percentages. These similarities between the large- and small-enrollment sections support the use of the small-enrollment class as a testing ground for a cooperative learning strategy that can, with modifications, be used in a large-enrollment section. The remainder of the methods and discussion will focus on the small-enrollment section of the course.

In the small-enrollment section, lectures, demonstrations, and in-class activities were given for about four days covering three to four textbook chapters. The fifth day, always on a Monday to give students the weekend to prepare, was a term examination that comprised short essays and short-answer questions. There were three term examinations given and a different-format comprehensive final examination, both described in more detail below.

The three term examinations, each worth a maximum of 100 points, had five multipart, equally weighted questions (exams 1 and 2) or ten equally weighted questions (exam 3). Questions required the students to remember definitions of terms (knowledge), identify the most important aspects of geologic processes (analysis), predict outcomes in one region based on a disaster episode in a different region (application), use maps to identify tectonic settings of current events (synthesis), illustrate (by drawing) geologic processes (knowledge), and perform simple calculations relevant to earth processes that result in natural disasters (application). The final comprehensive examination consisted of 97 multiple-choice and 18 true-false questions. Previously tested material questions composed 75% of the final examination, and new material questions composed 25% of the examination. The questions on the final examination made demands on students' knowledge similar to demands on the term examinations, with the exception of making drawings. Because the final examination was a different format from the term examinations, students were given a practice test consisting of 22 multiple-choice and four true-false questions as a downloadable document a week before the final examination. In addition, during the term, students were given lecture notes for each book chapter that included a section entitled "What to Learn," as a study aid.

For each of the three term examinations, the students took the examination individually and turned in their papers after a set amount of time had elapsed. After a short break, the instructor assigned students to groups and gave each student a blank examination, the same examination they had just completed individually. For the first term examination, groups were assigned after polling them for their major area of study. A group consisted of up to two students from each of the following general areas: (1) science/mathematics/engineering, (2) social sciences, (3) humanities, (4) business, and (5) fine arts. For the second and third examinations, a similar tactic was employed, with the additional discriminant such that groups also consisted of individuals with a spectrum of success on previous term examinations.

For the group examination, the same amount of time for completion was allowed for the individually completed examination, and students were permitted to discuss the answers to the questions with their group members only. They were also told that each student had to be responsible for at least one question (examinations 1 and 2) or two questions (examination 3), and the group members would decide who was the "leader" for each of the questions. Because one group had to consist of four people and there were either five or ten questions to be answered, in that group, only, one or two people had to be responsible for more questions; their grade was the average of their performance on the questions for which they were responsible. Groups were assigned so that students were in a group of four only once. Oral instructions given before the exams began emphasized that "being responsible" for a question meant leading the discussion about the question, but not necessarily knowing the correct answer. Students were also told that each group member was required to answer all the questions on the examination, but that only one examination from each group (chosen randomly) would be graded and that grade would be given to all group members. The weighting of the three parts of the examination (individual, group, question leader) was designed to motivate students to do well in all three aspects of the examination. Before the first examination, a handout was given to students and examples were described during class time to show that participating in the group examination would be to their advantage, and that if a student did well on the individual examination, they would not necessarily lose their high letter grade if their group got a lower score. Our method applies the jigsaw technique in the context of the tiered exam because it requires students to take the same examination twice (once individually and then in a small group) while each student is responsible for leading group discussion for at least one group examination question.

RESULTS

Almost all students received a higher group examination score than individual term examination score, for all three term examinations (Fig. 3). For only four out of the 57 paired scores did the score on the group examination decrease by more than 5% compared to the individual examination score.

We predicted that the students' performance on the final examination on the questions that related directly to questions on the term examinations (previously tested material) would show positive results as a result of the group examination. The results show that students earning the lowest scores on the individual midterm examinations improved their correct-answer percentage on the final examination questions covering previously tested material [plot above the 1:1 line, Fig. 4(a)]. When the results are separated into groups of those with individual-term-examination scores of less than 60%, 60%–69%, 70%–79%, 80%–89%, and greater than 89%, it is very clear that those with the lowest scores benefited the most from the group examination (Fig. 5). Those with individual-term-examination scores of less than 60% scored an average of 15% higher on the final examination questions covering previously tested material (Fig. 5). As individual-term-examination scores increased, the average difference between the term-examination score and the same-topic questions on the final examination decreased (Fig. 5).

Although it is difficult to assess improvement in learning, we chose to compare achievement on questions covering new material on the final examination (material covered in the last section of the term that had not been previously tested) with achievement on individual term

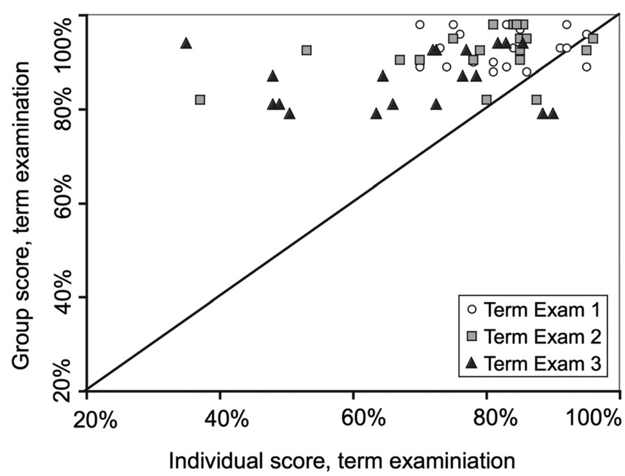

FIGURE 3: Most students earned higher scores on group examinations than individual examinations.

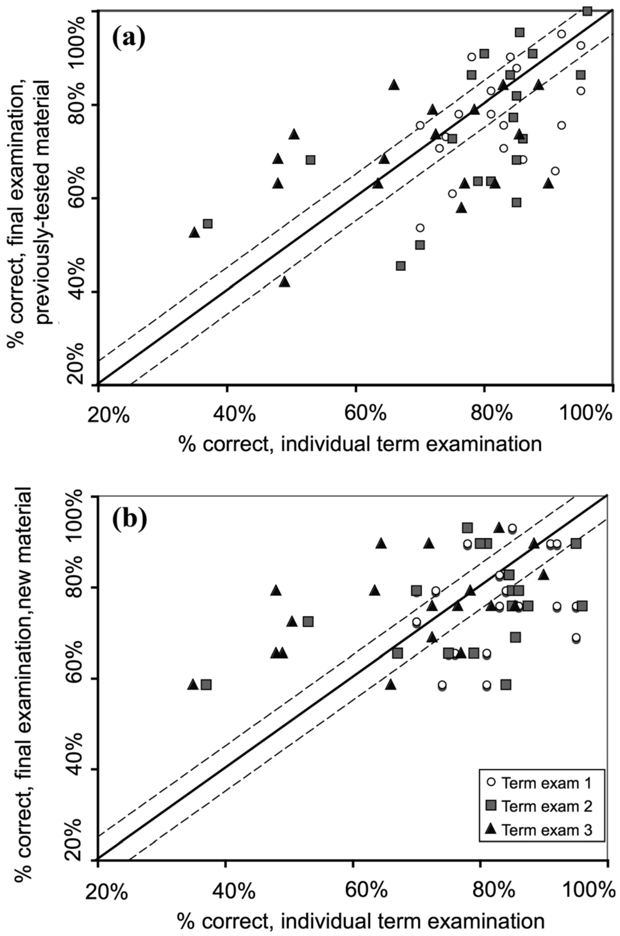

FIGURE 4: Comparison of performance on term examinations by individuals and final examination, separated according to whether questions were previously tested or new material. (a) Percentage correct answers on individual term examination vs percentage correct answers on final examination questions on material previously assessed in the term examinations. Solid line is 1:1; dashed lines are ±5% of 1:1 line. (b) Percentage correct answers on individual term examination vs percentage correct answers on final examination questions on new material (not previously assessed in the term examinations). Solid line is 1:1; dashed lines are ±5% ± 5% of 1:1 line. Symbols same as Fig. 4(a).

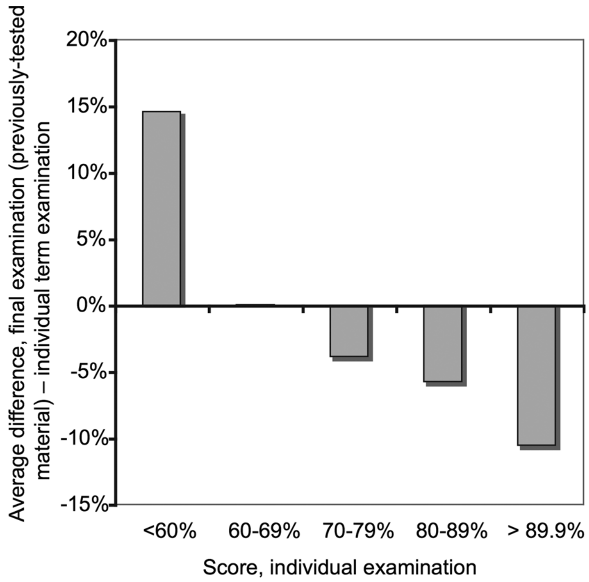

FIGURE 5: On average, the students earning the lowest scores on term examinations had the largest positive difference in score on the previously tested material on the final examination. There are 57 data pairs (three term exams, 19 students) that compare the term examination score with only those questions on the final examination covering the same topics as each term examination. The average of the difference between the two scores is shown here, categorized by score on the individual term examination.

examinations. When considering the new material on the final examination, the students with the lowest individual-term-examination scores also achieved a higher correct-answer percentage on the final examination questions covering new material than the total correct-answer percentage on the individual term exams [data plotting above the 1:1 line, Fig. 4(b)]. For the other students, about as many students scored higher as scored lower on that part of the final examination as they did on the individual-term-examinations.

In an anonymous poll given midterm, more than 70% of students agreed that they had learned from the group examination format. Half of the students said, if given the choice, they would choose the group examination over an individual examination, 36% said they preferred the individual examination, and 14% did not have a clear opinion either way. The university prescribes the format of the anonymous course survey given at the end of the term, so, instead of repeating the midsemester survey, students were asked to give their opinion of the group-examination format in their own words, if they wished. Even though the opinion poll format was different at the end of the semester (volunteered comments) than at the midpoint (specific questions), the results are that, at the end of the semester, more students favored the group-examination format (32% with purely positive comments) than did not (5% purely negative comments).

DISCUSSION

Students felt they learned from the group-examination format, although many expressed that the group examination took too long. In general, students who benefited the most from the group examination, if examination scores are a measure of that benefit, were the lowest achieving students. These students clearly improved their score not only on previously tested material, but also on new material. Whether the success of the low-scoring students is due to the peer teaching during group examinations or due to extra motivation to pass the course is ambiguous without a control group. However, the clear success of the low-achieving students on the previously tested material on the final examination suggests that, even if the benefit was not because of the format of the term examinations, the benefit might have resulted from increased incentive from their better success on the group-term examination, making them believe that they could be successful on the final examination.

In this pilot of the modification of the jigsaw technique applied to the tiered exam method, there is additional grading required compared to the traditional individual final examination format because the individual examinations are graded and one examination from each group is graded. If assigned groups consist of five students, this means an increase of 20% in the amount of grading that must occur. In addition, for this study, the format for the three term examinations was short- to medium-length essay questions. These questions included some that required student to draw diagrams in addition to writing out word answers. This question format is difficult to use in large classes because the time spent grading becomes unreasonable. A promising modification to this format for large classes is to administer an individual examination that could be machine-graded, and then administer a group examination with selected, similar-material, essay questions. With a well-written rubric, these group examinations, which would number ~20% of the size of the class (20 group assessments for a class of 100 students), could be graded in a reasonable amount of time. If learning by low-achievement students can be significantly improved by this method, whether the reason is because of interaction with other students or confidence building because of better success on examinations, it is a worthwhile endeavor.

CONCLUSIONS

We applied the jigsaw technique in the context of the tiered exam to allow students in a small-enrollment section, after completing an examination on their own, to discuss examination questions in groups designated by the instructor. In addition, each student was responsible for leading the discussion during group examination on at least one question, decided upon by the group members. Grades were assigned for individual performance, the performance of the group (same grade for all group members, based on grading on one member's examination paper), and the performance on the leader's question(s). These grades were weighted so that students were motivated to participate in the group and do well on the question(s) on which they were the discussion leader.

Scores on group examinations were usually higher than on individual examinations, and most students liked the group-examination format. The learning gained in the term examinations was mostly restricted to the lowest achieving students, when student performance on the final examination is used to assess learning. The group examination format benefited the lowest-achieving students the most, both for questions on previously tested material and for new material.

The assessment method as tested here, with essay-type questions, is most applicable to small classes because of the increased time spent grading by the instructor(s). To use the group-examination format reported in this paper in a large-population course, modifications could be made to administer a machine-gradable assessment as the individual examination and a person-graded examination for the group examination. The fact that the majority of students liked the group examination format shows both a willingness to learn more and a willingness to share one's knowledge, and it approaches more closely the workplace team, an environment into which we hope students will transfer soon after graduation, and for which we hope students will be well-prepared.

REFERENCES

Ben-Ari, R., and Kedem-Friedrich, P., 2000, Restructuring heterogeneous classes for cognitive development: Social interactive perspective: Instructional Science, v. 28, p. 153–167.

Blumenfeld, P.C., Marx, R.W., Soloway, E., and Krajcik, J. 1996, Learning with peers: From small group cooperation to collaborative communities: Educational Researcher, v. 25, p. 37–40.

Bykert-Kauffman, A.B., 1995, Using cooperative learning in college geology classes: Journal of Geoscience Education, v. 43, p. 309–316.

Crouch, C.H., and Mazur, E. 2001, Peer instruction: Ten years of experience and results: American Journal of Physics, v. 69, p. 970–977.

Fagen, A.P., Crouch, C.H., and Mazur, E., 2002, Peer instruction: Results from a range of classrooms: The Physics Teacher, v. 40, p. 206–209.

Kovac, J., 1999, Student active learning methods in general chemistry: Journal of Chemical Education, v. 76, p. 120–124.

Macdonald, R.H., and Korinek, L., 1995, Cooperative-learning activities in large entry-level geology courses: Journal of Geoscience Education, v. 43, p. 341–345.

McConnell, D.A., Steer, D.N., Owens, K.D., Knott, J.R., van Horn, S. Borowski, W., Dick, J., Foos, A., Malone, M., McGrew, H., Greer, L., and Heaney, P.J., 2006, Using conceptests to assess and improve student conceptual understanding in introductory geoscience courses: Journal of Geoscience Education, v. 54, p. 61–68.

Mazur, E., 1997, Peer instruction: A user's manual: Upper Saddle River, NJ, Prentice-Hall, p. 253.

Pear, J.J., and Crone-Todd, D.E., 2002, A social constructivist approach to computer-mediated instruction: Computers & Education, v. 38, p. 221–231.

Sheldon, A., 2006, Improve student learning through tiered exams: <http://serc.carleton.edu/NAGTWorkshops/hydrogeo/activities/9925.html>, 9 Mar. 2010.

Tewksbury, B.J., 1995, Specific strategies for using the "Jigsaw" Technique for working in groups in non-lecture-based courses: Journal of Geoscience Education, v. 43, p. 322–326.

Tewksbury, B.J., and Macdonald, R.H., 2005. Designing effective and innovative courses: <http://serc.carleton.edu/NAGTWorkshops/coursedesign/tutorial/assessment.html>, 9 Mar. 2010.

Tobin, K., 1993, The practice of constructivism in science education: Hillsdale, NJ, Lawrence Erlbaum Associates, p. 360.

Vygotsky, L., 1978, Mind in society: Development of higher psychological processes: London, Harvard University Press, p. 159.

Yuretich, R.F., Khan, S., Leckie, R.M., and Clement, J.J., 2001, Active-learning methods to improve student performance and scientific interest in a large introductory oceanography course: Journal of Geoscience Education, v. 49, p. 111–119.